DOCUMENT RESUME

ED 143 276 HE 009 182

AUTHOR Gammell, Linda
TITLE Student Ratings, Other Measures Help Assess Teaching Impact. Comment, Number 31.
INSTITUTION Minnesota Univ., Minneapolis. Center for Educational Development.
PUB DATE May 77
NOTE 7p.
AVAILABLE FROM Center for Educational Development, 317 Walter Library, University of Minnesota, Minneapolis, Minn. 55455

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Achievement; Classroom Observation Techniques; *College Faculty; Evaluation Criteria; Evaluation Methods; *Faculty Evaluation; *Higher Education; Literature Reviews; Peer Evaluation; Rating Scales; Self Evaluation; Student Evaluation of Teacher Performance; *Teaching Quality; *Teaching Skills

ABSTRACT

Evaluations of faculty instructional ability have long been used to assist in promotion, tenure, and hiring decisions or to help improve individual teaching performance. Internal and external pressures for increased accountability, together with improvements in evaluation methods, have resulted in increased emphasis on evaluation of instruction. Among the information sources currently used are measures of student achievement, classroom observations, and ratings by colleagues, administrators, or students. Measures of student learning as direct indicators of a faculty member's teaching ability are not always reliable, but do give some important information. Classroom observation can take a number of forms, but there is disagreement about the meaning and use of each. Student evaluations of courses and instructors can address many dimensions of teaching: teaching skill, rapport, interaction, feedback, organization and planning of the course, amount and difficulty of work, perceived value of the course, and perceived achievement. However, aspects of instructor performance measured by student ratings may be quite separate from the factors that result in effective teaching. Faculty self-ratings, peer ratings, and student ratings as measures of effectiveness generally agree in pattern; however, self-ratings and peer ratings tend to be higher than student ratings. (Author/MSE)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED143276

comment

CENTER FOR EDUCATIONAL DEVELOPMENT • UNIVERSITY OF MINNESOTA
Number 31 May, 1977

Student ratings, other measures help assess teaching impact

Evaluations of faculty instructional ability have long been used to inform promotion, tenure, and hiring decisions or to help improve individual teaching performance. Internal and external pressures for increased accountability together with improvements in evaluation methods have resulted in increased emphasis on evaluation of instruction. Among the information sources currently used are measures of student achievement; classroom observations; and ratings by colleagues, administrators, or students. This issue of Comment *looks briefly at each of these and then compares student ratings with other means of evaluating teaching.*

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

H009182

Student learning: one measure of teaching ability

The most basic measure of the effectiveness of a teaching program is the amount students learn that can be attributed to the program. Grades, course examinations, and other achievement measures all may be used to get some idea of the outcome of teaching activities. However, most who are familiar with this type of evaluation agree that improvement is needed in the design, use, and interpretation of achievement measures as indicators of teaching effectiveness.

One difficulty is that teaching and learning can take place somewhat independently of one another. A student may learn despite a poorly organized course and an ineffective instructor, or may fail (or refuse) to learn despite an excellent instructional program.

Another problem is that achievement scores are seldom reported in terms of gain, as with the use of pre- and post-tests. Thus attempts to measure the impact of a particular course and instructor may reflect to some unknown extent the preparation level at which students entered the course.

In addition, achievement scores are not often scaled against objectives for a given course, i.e., how students did in comparison with what they were expected to do. Further, achievement scores can be affected by other factors outside the direct control of the instructor, such as student motivation or the nature and content of the course.

Finally, a student may gain more in the classroom than is typically measured by achievement tests. For example, the teacher may have effected a change in student attitude, inspired the student to take more courses in a particular field, given the student a "feel" for an area of study or a set of skills useful for further learning in that area—all potential effects of good teaching.

All of these factors make it difficult to interpret measures of student learning as direct indicators of a faculty member's teaching ability. Nevertheless, such measures do give us some important information, and it is to be hoped that work will continue in an effort to improve their scope and usefulness.

Comment is the periodic publication of the University of Minnesota Center for Educational Development. James Werntz, Director. Gordon Wm. Kingston, Assistant Director. Andrew Ahlgren, Associate Director for Precollege Development. Anne Wirt, Research Associate. *Comment* staff: Nancy A. Peterson, Editor. Linda Gammell, Writer.

Teaching activities can be observed

Another way to seek knowledge about teaching performance is to observe the teaching process. A number of methods can be used to review an instructor's teaching activities. Kenneth Eble (1971) outlines some of them:

Classroom visitation. Though individual faculty members in many institutions may occasionally visit one another's classes, Eble believes few institutions have formal, systematic visitation procedures or a strong tradition of classroom visitation. Many attempts to institute the practice, or to continue it where it has existed, have met with resistance from faculty members, Eble reports, at least partly because they consider it intrusive and time-consuming.

It should be noted that these are Eble's views, and, further, that there are some long-standing classroom visitation programs at the University of Minnesota.

Although Eble does not say so, critics of the method also question its validity, believing that the presence of visitors in the classroom may create an artificial atmosphere and distort—for better or worse—the instructor's performance.

Self-appraisal. Self-evaluation in which an instructor clearly sets forth objectives and analyzes how those objectives might be achieved can be "a brilliant contribution to our knowledge about teaching," Eble believes. Some institutions ask for self-appraisals along with other forms of evaluation such as student questionnaires or peer review.

Course materials and data. Eble suggests that course syllabi and examinations may already be looked at by the department head either in passing or at the time of formal review. Other available measures of a faculty member's teaching activities include the number of student credit hours handled, kinds of courses taught, enrollments in elective classes, grade distributions, student performance on exams, and extent of supervision of independent study. Eble adds, however, that it is difficult to find agreement about the meaning and use of these measures.

Recording devices. Use of videotape equipment or, in some instances, audio equipment, affords an opportunity to observe and evaluate a teaching performance without the observer being present at the time. Eble suggests that this might reduce the "threatening or coercive aspects" of the process. Further, he suggests that the increasing use of videotaping and evaluation of graduate student teaching assistants and, often, beginning instructors, may lead to wider use for all faculty members.

Team teaching. "Team teaching, ordinarily thought of as benefiting the student, may be of equal benefit to the teacher," Eble says. "The chance to observe other teachers in action and to benefit from being observed is almost ideally present in team teaching." The reason: team teaching offers direct observation without alien intrusion into the classroom.

Students evaluate courses, instructors

Another way to gain information about the quality of an instructional program is to ask students to evaluate their courses and instructors. Although widely collected, student ratings are one of the most controversial ways to evaluate instruction, particularly since ratings of instructors may affect personnel decisions made about them.

Questions are raised regarding (a) the validity and reliability of student ratings (what they measure and how well they measure it) and (b) the extent of their relationship to effective teaching. Among those active in the study and debate on such questions are two faculty members of the University of Minnesota.

Kenneth O. Doyle, Jr., research associate at the University Measurement Services Center and author of a book entitled *Student Evaluation of Instruction* (1975) believes that student ratings are valid and reliable—though by no means perfect—indicators of teaching quality.

Paul W. Fox, professor of psychology, is writing a series of papers on the relationship between student ratings of instruction and measures of student achievement. As will be seen, Fox reports that the aspects of instructor performance measured by student ratings may be quite separate from the factors that result in effective teaching.

Questionnaires used by students to rate their instructors and courses generally contain a variety of items assessing student impressions of the knowledge and skills they gained in the course; the reading materials, workload, exams, and grading; and, of course, the performance of the instructor.

Students usually are asked to rate their instructors on many individual items such as clarity of presentation, enthusiasm, and use of class time. Several researchers have grouped related items into clusters expressing the dimensions on which students make judgments about their courses and instructors.

Illustrative of these dimensions is the set of six factors identified by researchers at the University of Michigan and described by James Kulik (1976).

Teaching skill. Items grouped into this category focus heavily on the student's opinion of the instructor's expositional skills. Typical items include "The teacher put across materials in an interesting way," and "Gave explanations that were clearly to the point." Other items in this category ask about the instructor's enthusiasm, knowledge of the subject, sensitivity to student difficulties, and the like. Teaching skill, according to several researchers, is the factor most closely identified with overall teaching performance.

Rapport. Items in this category measure how strongly the instructor is perceived as being friendly, democratic, open, and concerned. Sample items: "The teacher listened to students' questions and was willing to help," and "He explained the reasons for his criticism." At least one researcher has used a "warmth" vs. "coldness" rating scale.

Interaction. This category measures how much a teacher encourages student contributions in the classroom. Individual items might ask about the amount of classroom discussion, or students' freedom to disagree and ask questions.

Feedback. This set of items reflects perceptions of the degree to which the instructor encourages students, apprises them of their progress, and suggests ways they can improve.

Organization and planning of the course. This category assesses student opinion of how well the instructor has planned the course, keeps it on schedule, and sets out clear objectives and expectations.

Amount and difficulty of the work expected. Students are asked to respond to such items as those covering the amount of work required, amount of material covered, and the standards set for students.

Fox and Edwall (1976) have also identified two additional sets of items used by some researchers:

Perceived value of the course. Several questionnaires have included rating scales in which students are asked to assess the overall value of the course.

Perceived achievement. Included in this category are items asking students to assess the amount they have learned in the course.

Student, peer, self ratings similar in pattern

How do student ratings compare with other sources of evaluation information? Some attempts have been made to compare student ratings with those by colleagues and administrators and with instructors' self-ratings. According to Doyle, these comparisons have shown that colleagues tend to be a bit lenient, rating a peer somewhat higher than students would. Ratings by administrators are very close to those of colleagues.

Self-ratings are reported to be questionable assessments of one's teaching effectiveness. James Kulik (1974) explains that the problem may be a matter of numbers. "Ratings by students, colleagues, and administrators are composite ratings. . . . Idiosyncrasies of viewpoint cancel each other out. The severe judgment by one student may be balanced by a more lenient rating from another and, thus, a tempered judgment emerges from disparate views. Self-ratings, on the other hand, are not composites but ratings from a single individual. If the self-evaluating teacher is dull but generous, a poor teacher will get a favorable review; if he is brilliant and self-effacing, a good teacher will get a poor rating."

Doyle reports that self-ratings, peer ratings, and student ratings generally agree in pattern—that is, they tend to rank an instructor's characteristics in the same order of strength. However, instructors usually rate themselves higher than their students rate them.

Ratings vs. learning: comparison stirs debate

Many researchers have attempted to compare student ratings of instruction with measures of student achievement, using achievement measures as a criterion against which to test the validity of student ratings. Fox and Edwall (1976) recently reviewed these studies. In an increasingly detailed series of analyses, they examined the reported relationships between achievement measures and student ratings of instruction. In effect, they were asking whether the amount students learned, as measured by the many kinds of achievements tests used in these studies, had a relationship to how the students rated their instructors. If no such relationship were found, they suggested, the factors measured by student ratings, whatever else their value, may not be the factors that determine whether a faculty member is an effective instructor.

Fox and Edwall found a wide range of correlations, from moderately negative to strongly positive, many of which were difficult to take very seriously. "Overall, it appears that only a small number of studies have obtained reliable correlations between evaluative factors and measures of achievement," they stated.

They pointed out that attempts to study the relationship between learning and evaluation may be affected by (a) measurement error resulting from the variety of achievement measures and rating instruments used in the original studies, (b) the disparity in research design and statistical methods used, and (c) the number of additional factors that may influence evaluations, such as subject matter area, course level, whether the course is elected or required, class size, and instructor level.

"In our opinion," Fox and Edwall concluded, "present evidence on the nature of the relationship between teacher ratings and learning is rather weak and equivocal. At each level of analysis, no single clear pattern has emerged, except perhaps the general tendency of the obtained relationships to be positive. On balance, however, the average magnitude of the relations examined appears small and hence suspect as a valid measure of 'teacher effectiveness.'

"Student ratings inform us about student attitudes and perceptions of instruction, and this may be quite valuable information for educators to obtain. To learn about teacher effectiveness, however, at least as this is referenced by student achievement,

may very well demand quite different approaches from those embodied in student ratings."

Doyle does not believe that student learning is a legitimate criterion against which to test the validity of most evaluation items, whether those items are responded to by students, peers, or instructors themselves. He believes that to evaluate the validity of ratings in terms of student learning requires looking at each rated instructor characteristic separately and deciding how that characteristic should relate to student learning, if at all.

For example, he said, "there is little reason to expect that the clearer an instructor is, the more the students should learn. Too little clarity probably inhibits learning, but too much might result in superficiality or boredom." Similarly, he suggested, "some threshhold amount of rapport might be all that is necessary. Thus there is no reason to look for a simple positive correlation between learning and ratings of rapport.

"On the other hand," Doyle continued, "ratings on some items probably should correlate with student learning—ratings of overall teaching ability, for example." He believes such correlations are consistent though usually "quite modest." Unlike Fox, Doyle tends to be satisfied that these existing correlations do signal a relationship between ratings and learning.

Noting that complex methodology is involved in studies of the relationship between ratings and learning, Doyle listed some of the factors that he said can reduce the size of the correlation found between the two: unreliability in either the ratings or the learning measures; departures from random sampling in the choice of subjects; restrictions on the range of data such as would arise from studying select groups of students or instructors; and differences in the measuring scales used (e.g., a continuous scale for the learning measure and a seven-point scale for the rating).

"Couple all of this with the limited content of many examinations and the difficulty of measuring student gains from beginning to end of a course, and it becomes really difficult to know what these studies are saying," Doyle concluded. "In short, while student learning may *in theory* be a legitimate criterion for some evaluation items, *in practice* studies of this sort aren't nearly so informative as they might seem."

Doyle believes that the best solution for the present is to make "good-faith judgments"—based, he said, on reason, experience, and what data are available—about which instructor characteristics are important to teaching in a particular context, and then to get evaluations from students "and anybody else" on those items. He stressed that the questions raised about the validity of student evaluations "apply with equal force to any other information that might be used to evaluate instruction, including tests and peer evaluations."

What do student ratings tell us?

Fox said he believes there are three principal questions that should be asked about student evaluations of instruction (or about any other form of faculty evaluation):

(1) ***Do they measure a faculty member's knowledge and competence in the subject matter area?*** Fox's answer is that there is no evidence to show that student ratings measure professional knowledge and competence, and some evidence that says they don't. Both Fox and Doyle suggest that peers are probably in a better position than students to evaluate subject matter competence.

(2) ***Do they measure teaching effectiveness?*** It is in their answer to this question that the differences between Fox and Doyle become most apparent. Fox says, "If effectiveness is taken to mean some measurable intellectual change brought about as a consequence of the instruction, the date provide precious little support for the view that ratings are valid measures of teaching effectiveness."

Doyle, by contrast, believes that while student ratings are "far from perfect," they are "usually pretty good measures of the quality of teaching, and still the best measure available."

Doyle's use of the general term "quality" in this context, rather than the more specific term "effectiveness," signals a slightly different meaning that shifts the argument. While "effectiveness" focuses on learning *outcomes*, "quality" can include judgments on aspects of the teaching *process*.

(3) ***Do they measure student opinion of the teaching process?*** Fox and Doyle agree that student evaluations are reasonably reliable and valid measures of student opinion. They appear to differ, however, on the *importance* of such opinion. The importance of student opinion as a predictor of learning outcomes has already been considered under "effectiveness" above. However, some aspects of the teaching process might be considered important regardless of whether they could ever be demonstrated to lead to measurable learning outcomes—either because they have purported learning outcomes that can never be measured or because they are believed to have intrinsic "humanistic" or "traditional" value. Such beliefs about what constitutes "quality" teaching would be appropriate for philosophical debate, but would not be subject to empirical research.

Appeal to faith or common sense or principles may seem tenuous in a discussion that began in a review of empirical research, but it is not easy to judge where the more useful truth may lie—it is, after all, on these tenuous bases that we conduct most of our human affairs.

Caution suggested in use of ratings

Whether student ratings are seen primarily as measures of student opinion, as in Fox's view, or as the best available indicators (however flawed) of teaching quality, as in Doyle's, caution must be exercised in their use, particularly when used in making decisions about tenure, salary, or promotion.

Evidence seems to indicate that ratings can be affected at least somewhat by such factors as class size, discipline, and course content. Obviously, an instructor assigned to a large, required course could be at a disadvantage in comparison with one who teaches a small elective seminar. Less dramatic instances, not so apparent to those reviewing the ratings, could be misleading.

Because of the difficulty of interpreting ratings, it is often suggested that institutions develop norms against which to compare an individual's ratings. "In the case of large courses consisting of many sections, one per teacher, such norms can be developed easily," James Kulik says. "It is more difficult to develop adequate norms for evaluating the hundreds of instructors who teach in non-sectioned courses. To evaluate such teachers, comprehensive, college-wide data on courses and students would have to be collected and used as the basis for norms for courses classified according to size, department, content, elected or required, etc." The problems are obvious.

Individual instructors also may use evaluations to improve their teaching, although researchers have found little evidence that they have done so. However, most studies have concentrated on a very short time span, usually comparing an instructor's ratings obtained midway through a course with those received at the end of the course. Instructors may need more time to incorporate information from ratings into their classroom performance, Kulik suggests. In addition, instructors most often receive their own ratings without additional materials to help them interpret and respond to those ratings.

Fox cautions that work he is currently doing suggests that average and higher ratings should be regarded only as very approximate indications of student approval level; relative standings within this overall range are very difficult to interpret. Consistently very low ratings, however, probably indicate that the situation needs to be looked into.

He believes that acceptance and use of ratings would increase if people were to perceive them as measures of student perceptions and opinions. He added that he is impatient with the amount of attention devoted to the "false issues" that arise when student ratings are seen as a definitive assessment of "teaching ability." This attention, he said, diverts from what he sees as the real work of developing better measures of learning and teaching effectiveness, and finding ways to improve instruction.

Units help faculty improve instruction

Evaluation of teaching continues to be a lively topic of study and discussion, and this brief review is not meant as a definitive description of either the overall issue or the positions of Fox and Doyle, which continue to evolve even as this issue is written and likely will be the subjects of further publication by each.

Faculty members at the University of Minnesota who wish information about their teaching and how to improve it may turn to a number of University-wide service units.

The Measurement Services Center provides assistance with developing and processing student evaluation questionnaires. Faculty members using questionnaires processed by the center receive an interpretive guide for use with the reports. The guide includes step-by-step suggestions for using the ratings, a bibliography on teaching and evaluation, and hints on improving teaching performance. The center also sponsors workshops on using evaluations to improve instruction, and currently is preparing more detailed and individualized materials intended to help faculty make fuller use of student ratings of instruction. The center is located at 9 Clarence Ave. S.E., Minneapolis, (612) 373-2263.

The Consulting Group on Instructional Design works through faculty seminars, consultation, and cooperative efforts with faculty to develop new approaches to instructional programs, often using new instructional techniques and technologies. The CGID is at 205 Elliott Hall, Minneapolis, (612) 373-5352.

Funds for projects aimed at improving instruction are available through three development grant programs administered through the Center for Educational Development: the University-wide Educational

References

Doyle, Kenneth O., Jr. 1975. *Student Evaluation of Instruction*. Lexington, MA: Lexington Books, D.C. Heath & Co.

Eble, Kenneth. 1971. *The recognition and evaluation of college teaching*. Salt Lake City: Project to Improve College Teaching.

Fox, P.W., and Edwall, G.E. 1976. *The validity and use of student evaluations of instruction: I. The relation between student ratings of faculty and student achievement*. Consulting Group for Instructional Design, University of Minnesota.

Kulik, James A., and Kulik, Chen-Lin C. 1974. Student ratings of instruction. *Teaching of Psychology*. 1 (December 1974):51-57.

Kulik, James A. 1974. Evaluation of teaching. *Memo to the faculty*. 53:1-5. Center for Research on Learning and Teaching, University of Michigan.

Development Program, the Small Grants Program of the all-University Council on Liberal Education, and the Media Production Fund. Information is available from 317 Walter Library, Minneapolis, (612) 373-4537.

In addition to these University-wide service units, a number of collegiate and campus units exist to help faculty members improve the quality of instruction at the University.

'Comment'

The University of Minnesota has in place two major and interrelated processes for improving its instructional programs. The more fundamental of the two is the process by which the faculty judges itself personally, for self-improvement, and collectively, for assigning the rewards and encouragements of salary and academic rank. Of nearly as great importance is the process that leads from program review and evaluation through development to improved operation and subsequent re-evaluation. Both processes require assessments from outside the University, whether from national and international peer groups or from supporting publics; both require assessments from inside the University from self, peers, colleagues, and students.

This issue of *Comment* has attempted to sharpen general understanding of the characteristics and uses of information elicited from students regarding the instructional programs of the University. Some interesting haziness is yet visible around the outer limits of our present understanding, suggesting needs for new information and new uses for existing information.

- Traditionally, measurements of student perception of the effectiveness of instruction have been used primarily for evaluation of program and faculty. To what extent could these measurements also be useful to students for their own program planning?
- Traditionally, the University has attempted to assess the academic achievement of individual students in order to certify their learning. How can academic achievement of students also be used systematically as a general measure of the effectiveness of programs and faculty?
- Even after the final wisp of haze has evaporated from our understanding of the relationship between academic achievement and student perceptions of the quality of the education provided them, the question will persist: How should we balance the emphasis between these two important measures in our various uses of them?
- What is the significance of basing the assessment of the effectiveness of University programs on achievement in *courses* and student satisfaction with *courses*? So doing implies a compartmentalization that may obscure important elements of the development of the human intellect and thus may impede the improvement of University instructional programs.

While awaiting the haze to clear, we will continue to need solid information about both the academic achievement of students and their perception of the quality of their educational experiences at the University; both are essential to the processes of improving University programs. The information from students is complex and important. Its complexity should make us skeptical and cautious as we employ it. Its importance is a challenge to us to improve its quality, relevance, and use.

James H. Werntz, Jr., Director
Center for Educational Development

comment
Center for Educational Development
317 Walter Library
117 Pleasant Street S.E.
University of Minnesota
Minneapolis, Minnesota 55455